The Annals of Statistics
1985, Vol. 13, No. 1, 163–177

ESTIMATING A DISTRIBUTION FUNCTION WITH TRUNCATED DATA

BY MICHAEL WOODROOFE[1,2]

The University of Michigan and Rutgers University

Let $\mathscr{P}$ be a finite population with $N \geq 1$ elements; for each $e \in \mathscr{P}$, let X_e and Y_e be independent, positive random variables with unknown distribution functions F and G; and suppose that the pairs (X_e, Y_e) are i.i.d. We consider the problem of estimating F, G, and N when the data consist of those pairs (X_e, Y_e) for which $e \in \mathscr{P}$ and $Y_e \leq X_e$. The nonparametric maximum likelihood estimators (MLEs) of F and G are described; and their asymptotic properties as $N \to \infty$ are derived. It is shown that the MLEs are consistent against pairs (F, G) for which F and G are continuous, $G^{-1}(0) \leq F^{-1}(0)$, and $G^{-1}(1) \leq F^{-1}(1)$. $\sqrt{N} \times$ estimation error for F converges in distribution to a Gaussian process if $\int_0^\infty (1/G)\, dF < \infty$, but may fail to converge if this integral is infinite.

1. Introduction. Consider a finite population $\mathscr{P}$ whose size N is large, but otherwise unknown. For each element $e \in \mathscr{P}$, let X_e and Y_e denote independent, positive random variables with distribution functions F and G, say; and suppose that (X_e, Y_e), $e \in \mathscr{P}$, are i.i.d., as (X, Y), say. Finally, suppose that one observes (only) those pairs (X_e, Y_e) for which $Y_e \leq X_e$, but not the labels $e \in \mathscr{P}$. The problem considered is that of estimating F, G, and N. Nonparametric maximum likelihood estimators (MLEs) of F and G, described in (8) and (9) below, have been derived by several authors, listed below, from different perspectives. Here the asymptotic properties of the estimators are studied, and still another derivation suggested.

This model arises in astronomy. The absolute and apparent luminosities of an astronomical object are defined to be its brightness at a fixed distance and as observed on earth; and magnitude is defined to be the negative logarithm of luminosity. In some models, the redshift z and the absolute magnitude M of astronomical objects are assumed to be independent random variables which are related to the apparent magnitude m by the equation

$$m = f(z) + M, \tag{1}$$

where f is a known function, or at least a nearly known one. For example, Hubble's Law specifies that $f(z) \approx 5 \log z$, and Segal's Chronometric Theory specifies that $f(z) \approx (5/2)\log[z/(1 + z)]$. See Segal (1975). Of course, one can

Received June 1983; revised March 1984.

[1] Presented at the Jack Kiefer-Jacob Wolfowitz Memorial Statistical Research Conference; dedicated to their memories.

[2] Research supported by the National Science Foundation under MSC-8101897.

AMS 1980 *subject classifications.* Primary 62F20; secondary 62G05.

Key words and phrases. Nonparametric, maximum likelihood estimation, consistency, asymptotic distributions.

only detect objects which are sufficiently bright, say $m \le m^*$. Then, letting $X = \exp[-f(z)]$ and $Y = \exp[M - m^*]$ yields the model described above.

In other applications, the X_e may be the sizes of hidden objects for which one searches for one unit of time and $T_e = Y_e/X_e$ might be the time at which one would find the object e, if the search were continued indefinitely. Then the conditional probability of finding object e given X_e is $G(X_e)$, an unknown but increasing function of X_e. For example, Barouch and Kaufman (1975) have described models for exploring for petroleum reserves in which the probability of finding a given pool is proportional to the pool's size. Letting X denote a pool's size and T denote the time at which it would be found in an infinite search yields a model which is closely related to Barouch and Kaufman's (1975).

Starr (1974), Starr, Wardrop, and Woodroofe (1976), and Kramer (1983) have considered a class of optimal stopping problems in which one searches for hidden objects and receives a reward depending on the objects found, say the sum of their sizes, less a cost of sampling. Assuming a known stochastic model and certain other conditions, these authors obtain explicit solutions to the optimal stopping problem. In addition, they propose adaptive procedures for use when the total number of objects N is unknown. The estimators studied here may allow implementation of adaptive procedures in which other quantities, like F, are estimated sequentially.

Nonparametric MLEs of F and G were derived by Lynden-Bell (1971), who described another application to astronomy. See also Jackson (1974). Nicoll and Segal (1980) derive the MLEs for grouped data; and Bhattacharya, Chernoff, and Yang (1983) derived MLEs from a conditional likelihood function of certain counts, given the observed X-values. The latter paper also computes the information matrix for its model. Bhattacharya, et al. (1983) construct nonparametric estimators of regression parameters in models like (1), and show asymptotic normality of estimation error, properly normalized; and Bhattacharya (1983) considers the asymptotic distribution of a goodness of fit statistic with a view towards testing hypotheses about regression parameters. None of these papers give conditions for the consistency and asymptotic normality of the MLEs of F and G, however.

Here asymptotic properties of these estimators are studied as $N \to \infty$. In Section 2, the conditional distributions of X and Y given $Y \le X$ are related to the unconditional distributions F and G. The estimators are described in Section 3. Section 4 considers consistency; if F and G are continuous and if the lower and upper endpoints of the convex support of G are individually less than or equal to those of F, then the estimators converge to the true distribution functions F and G in probability as $N \to \infty$. Sections 5 and 6 consider normalized estimation error for the distribution functions. Here $\sqrt{N} \times$ estimation error for F converges in distribution to a Gaussian process if $\int_0^\infty (1/G)\, dF < \infty$; but the asymptotic variance may be infinite if this integral diverges.

There is some similarity between the estimators studied here and the estimator of Kaplan and Meier (1958), and hence with the asymptotic results of Breslow and Crowley (1974). There are also differences. The Kaplan Meier estimator would be appropriate if $X_i \wedge Y_i = \min(X_i, Y_i)$ and $\delta_i = I\{X_i \le Y_i\}$ were observed for $1 \le i \le N$; here both X_i and Y_i are observed if $Y_i \le X_i$, and nothing is observed otherwise. In terms of the asymptotic distributions, this difference leads to the possibility of an infinite variance for $\sqrt{N} \times$ estimation error.

There is also some similarity with recent results of Vardi (1982a, 1982b). He considers generalizations of our model when G is known, and obtains both nonparametric MLEs and asymptotic distributions.

2. A Transformation.

Let X and Y denote independent, positive random variables with distribution functions F and G, taken to be continuous from the right. Let H_* denote the joint distribution function of X and Y given $Y \le X$; and let F_* and G_* denote the marginal distribution functions of X and Y given $Y \le X$. Thus,

$$H_*(x, y) = \alpha^{-1} \int_0^x G(y \wedge z)\, dF(z),$$

$$F_*(x) = H_*(x, \infty) \quad \text{and} \quad G_*(y) = H_*(\infty, y), \quad 0 \le x, y < \infty, \tag{2}$$

where $\alpha = \int_0^\infty G(z)\, dF(z) = \int_0^\infty [1 - F(z-)]\, dG(z)$ is assumed to be positive. Here $y \wedge z$ denotes the minimum of y and z for $0 \le y, z < \infty$; $F(z-) = P\{X < z\}$ for $z \ge 0$; and $\int_a^b = \int_{(a,b]}$ for $0 \le a < b \le \infty$. There is little hope of finding consistent estimators of F and G from the data described in the introduction, unless F_* and G_* determine F and G. So, this question is investigated first.

If K is any distribution function on $[0, \infty)$, let

$$a_K = \inf\{z > 0: K(z) > 0\} \ge 0$$

and

$$b_K = \sup\{z > 0: K(z) < 1\} \le \infty,$$

so that (a_K, b_K) is the interior of the convex support of K. Then $\alpha > 0$ in (2) if $a_G < b_F$, and $\alpha = 0$ unless $a_G \le b_F$. If $\alpha > 0$ and if F_* and G_* are related to F and G by (2), then $a_{F_*} = \max\{a_F, a_G\}$, $b_{F_*} = b_F$, $a_{G_*} = a_G$, and $b_{G_*} = \min\{b_F, b_G\}$. In addition, it is convenient to have the following notation: let

$$\mathscr{K} = \{(F, G): F(0) = 0 = G(0),\ \alpha(F, G) > 0\},$$

$$\mathscr{K}_0 = \{(F, G) \in \mathscr{K}: a_G \le a_F,\ b_G \le b_F\},$$

$$T(F, G) = H_*, \quad (F, G) \in \mathscr{K}.$$

Lemma 1. (i) *Let* $(F, G) \in \mathscr{K}$ *and let* F_0 *and* G_0 *denote the conditional distributions of* X *and* Y *given* $X \ge a_G$ *and* $Y \le b_F$. *Then* $(F_0, G_0) \in \mathscr{K}_0$ *and* $T(F_0, G_0) = T(F, G)$;

(ii) $T(\mathscr{K}) = T(\mathscr{K}_0)$.

Proof. Since $Y \le X$ implies $X \ge a_G$ and $Y \le b_F$ w.p.1, $T(F, G) = T(F_0, G_0)$. To see that $a_{G_0} \le a_{F_0}$, observe that $a_{G_0} = a_G$, since $(F, G) \in \mathscr{K}$, and that $a_{F_0} = \max(a_F, a_G) \ge a_G = a_{G_0}$. A similar argument shows that $b_{G_0} \le b_{F_0}$ to complete the proof of (i). Assertion (ii) then follows since $\mathscr{K}_0 \subset \mathscr{K}$.

Recall that the cumulative hazard function of a distribution function F (with $F(0) = 0$) is defined by

$$\Lambda(x) = \int_0^x dF(z)/[1 - F(z-)], \quad 0 \le x < \infty.$$

The cumulative hazard function Λ uniquely determines the distribution F by the following algorithm; let D denote the set of x for which $0 \le x < b_F$ and $\lambda(x) = \Lambda(x) - \Lambda(x-) > 0$; then

$$1 - F(x) = \{\textstyle\prod_{z\in D, z\le x}[1 - \lambda(z)]\}\exp[-\Lambda_c(x)], \quad 0 \le x < b_F, \tag{3}$$

where $\Lambda_c(x) = \Lambda(x) - \sum_{z\in D, z\le x} \lambda(z)$, $0 \le x < b_F$.

THEOREM 1. *Suppose that $H_* \in T(\mathscr{H})$. Then there is a unique pair $(F, G) \in \mathscr{H}_0$ for which $T(F, G) = H_*$. Here the pair (F, G) is determined by the conditions*

$$\Lambda(x) = \int_0^x dF_*(z)/C(z), \quad 0 \le x < \infty, \tag{4}$$

and

$$\int_y^\infty dG(z)/G(z) = \int_y^\infty dG_*(z)/C(z), \quad 0 \le y < \infty,$$

where

$$C(z) = G_*(z) - F_*(z-), \quad 0 \le z < \infty.$$

PROOF. By the lemma, there is at least one pair $(F, G) \in \mathscr{H}_0$ for which $T(F, G) = H_*$. It is shown below that (4) holds for any such pair, and it then follows that there is only one such pair, by (3) applied to F and G_1, where $G_1(z) = 1 - G(1/z-)$, $z > 0$. The proof of (4) depends on the simple identity $C(z) = \alpha^{-1}G(z)[1 - F(z-)]$ for $z \ge 0$, which may be derived as follows:

$$\begin{aligned}\alpha C(z) &= P\{Y \le X,\ Y \le z\} - P\{Y \le X,\ X < z\}\\ &= P\{Y \le X,\ Y \le z \le X\}\\ &= P\{Y \le z\} - P\{X < z,\ Y \le z\} = G(z)[1 - F(z-)]\end{aligned}$$

for $0 \le z < \infty$. Since $a_G \le a_F$, it follows easily that

$$\begin{aligned}\int_0^x dF_*(z)/C(z) &= \int_{a_F-}^x G(z)\, dF(z)/\alpha C(z)\\ &= \int_{a_F-}^x dF(z)/[1 - F(z-)] = \Lambda(x)\end{aligned}$$

for all $x \ge a_F$; and both sides vanish for $x < a_F$. This establishes the first assertion in (4) and the second may be established similarly.

COROLLARY 1. *Let $(F, G) \in \mathscr{H}$ and let F_0 and G_0 be the conditional distributions of X and Y given $X \geq a_G$ and $Y \leq b_F$, as in Lemma 1. Then (F_0, G_0) is the only pair in $\mathscr{H}_0$ for which $T(F_0, G_0) = T(F, G)$.*

COROLLARY 2. *Let T_0 denote the restriction of T to $\mathscr{H}_0$. Then T_0 has an inverse function.*

PROOFS. Lemma 1 asserts that $(F_0, G_0) \in \mathscr{H}_0$ and $T(F_0, G_0) = T(F, G)$; and the theorem asserts that there is only one such pair. This establishes the first corollary. The second then follows, since $(F_0, G_0) = (F, G)$ when $(F, G) \in \mathscr{H}_0$.

REMARKS 1. The inversion formula of Theorem 1 uses only the marginal distributions of H_*.

2. Let $\mathscr{F}$ denote the class of all distribution functions on $[0, \infty)$. Endow $\mathscr{F}$ with its weak topology; endow $\mathscr{F} \times \mathscr{F}$ with the product topology; and endow $\mathscr{H}$, $\mathscr{H}_0$, and $T(\mathscr{H})$ with their relative topologies. Then T is easily seen to be continuous at all $(F, G) \in \mathscr{H}$ which have no common points of discontinuity. However, the inverse transformation to T_0 is not continuous. To see this let F and G be continuous distribution functions with support $[0, \infty)$; and let $G_n = (G + \delta_n)/2$, where δ_n denotes the point mass at n for $n \geq 1$. Then $T(F, G_n) \to T(F, G)$ as $n \to \infty$, but G_n does not converge to G.

3. Estimation. Now let F and G denote distribution functions for which $(F, G) \in \mathscr{H}$; let X and Y denote independent random variables with distribution functions F and G; and let $(X_1, Y_1), \cdots, (X_N, Y_N)$ be i.i.d. as (X, Y). As in the introduction, suppose that one observes only those pairs (X_i, Y_i) for which $i \leq N$ and $Y_i \leq X_i$. Suppose that there is at least one such pair, and let $(x_1, y_1), \cdots, (x_n, y_n)$ denote these pairs, so labeled that $(x_1, y_1), \cdots, (x_n, y_n)$ are conditionally i.i.d. given n.

To describe the estimators of F and G, let F_n^* and G_n^* denote the empirical distribution functions of $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$,

$$F_n^*(z) = (1/n)\ \#\{i \leq n: x_i \leq z\},$$
$$G_n^*(z) = (1/n)\ \#\{j \leq n: y_j \leq z\}, \quad 0 \leq z < \infty, \tag{5}$$

where $\# A$ denotes the cardinality of a set A. Thus, F_n^* and G_n^* estimate the conditional distribution functions F_* and G_*. Estimators of F and G may be constructed from F_n^* and G_n^* by using the inversion formula of Theorem 1. Let

$$C_n(z) = G_n^*(z) - F_n^*(z-), \quad 0 \leq z < \infty, \tag{6}$$

and observe that $C_n(x_i) \geq 1/n$ for all $i \leq n$. Then Theorem 1 suggests estimating the cumulative hazard function Λ by

$$\hat{\Lambda}_n(z) = \int_0^z dF_n^*(x)/C_n(x) = \textstyle\sum_{i:x_i \leq z} 1/nC_n(x_i), \quad 0 \leq z < \infty. \tag{7}$$

Observe that $\hat{\Lambda}_n$ is a step function with discontinuities (only) at $x_1, \cdots, x_n$. Thus, Equation (3) suggests estimating F by

$$\hat{F}_n(z) = 1 - \prod'_{i:x_i \le z} [1 - r(x_i)/nC_n(x_i)], \quad 0 \le z < \infty, \tag{8}$$

where $r(x_i) = \# \{k \le n: x_k = x_i\}$ for $1 \le i < n$, the product extends over distinct values of $x_1, \cdots, x_n$, and an empty product is to be interpreted as one. Of course, a similar construction is possible for the estimation of G. After some algebra, one is led to the estimator

$$\hat{G}_n(z) = \prod'_{j:y_j > z} [1 - s(y_j)/nC_n(y_j)], \quad 0 \le z < \infty, \tag{9}$$

where $s(y_j) = \# \{k \le n: y_k = y_j\}$ for $1 \le j \le n$.

The estimators $\hat{F}_n$ and $\hat{G}_n$ were derived by Lynden-Bell (1971). Suppose, for simplicity, that there are no ties among $x_1, \cdots, x_n, y_1, \cdots, y_n$ and consider estimating F and G by distributions which are supported by $\{x_1, \cdots, x_n\}$ and $\{y_1, \cdots, y_n\}$. For such distributions, the conditional likelihood function given n is

$$L_n = \alpha^{-n} p_1 \times \cdots \times p_n \times q_1 \times \cdots \times q_n,$$

where $p_1, \cdots, p_n$ and $q_1, \cdots, q_n$ are the masses assigned to $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$. This likelihood function may be maximized with respect to $p_1, \cdots, p_n$ and $q_1, \cdots, q_n$; and the estimators $\hat{F}_n$ and $\hat{G}_n$ result, provided that (10) below does not occur. Alternatively, one may show that F_n^* and G_n^* are the nonparametric, maximum likelihood estimators of F_* and G_* and then use the invariance properties of maximum likelihood estimators. The alternative derivation is not substantially simpler than the direct one, however.

The estimators $\hat{F}_n$ and $\hat{G}_n$ may be supported by proper subsets of $\{x_1, \cdots, x_n\}$ and $\{y_1, \cdots, y_n\}$. Let $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ and $y_{(1)} < \cdots < y_{(n)}$ denote the ordered values of $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$. If

$$nC_n[x_{(k)}] = 1, \quad \text{for some } k, \quad 1 \le k < n, \tag{10}$$

then

$$\hat{F}_n[x_{(k)}] = 1.$$

This a disturbing property of the estimators, since it may lead to unreasonable estimates. For example, it is possible to have $\hat{F}_n[x_{(1)}] = 1$. It is shown below that the probability of (10) approaches zero as $N \to \infty$, if F and G are continuous; but this will be of little comfort when (10) occurs.

The problems which result from (10) may be overcome in a simple, if ad hoc, manner. Let k_n be a nonincreasing function for which $k_n(x) > k_n[x_{(n)}] = 1/n$ for all $x < x_{(n)}$. If C_n is replaced by

$$C_n^{\#}(z) = \max\{C_n(z), k_n(z)\}, \quad 0 \le z \le x_{(n)},$$

in (9), then the resulting estimator $F_n^{\#}$ is not supported by any proper subset of $\{x_1, \cdots, x_n\}$. In fact, $1/nk_n[x_{(i)}]$ is the maximum proportion of the estimated probability $1 - F_n^{\#}[x_{(i)}-]$ which the experimenter is willing to assign to $x_{(i)}$ for $i = 1, \cdots, n$.

TABLE 1
Calculation of $\hat{F}_n$

k	$x_{(k)}$	y	$C_n(x_{(k)})$	$\hat{F}_n(x_{(k)})$	p_k
1	.3156	.0672	6	.1667	.1667
2	.3597	.0136	5	.3333	.1667
3	.4017	.0816	4	.5000	.1667
4	.4970	.4117	5	.6000	.1000
5	.5068	.2559	4	.7000	.1000
6	.6586	.1113	4	.7750	.0750
7	.7719	.5820	4	.83125	.05625
8	.7897	.4106	3	.88750	.05625
9	.8707	.0592	2	.94375	.05625
10	.9441	.7175	1	1.0000	.05625

The (x, y) pairs are listed in order of increasing x values; and $p_k = \hat{F}_n[x_{(k)}] - \hat{F}_n[x_{(k-1)}]$, $k = 1, \cdots, 10$. The sample average and MLE of the mean of F are $\bar{x} = .6116$ and $\hat{\mu} = .5192$.

It is especially interesting that one may estimate α, the probability that $Y \le X$, when one observes only those pairs (X_i, Y_i) for which $i \le N$ and $Y_i \le X_i$. The nonparametric maximum likelihood estimator of α is

$$\hat{\alpha}_n = \int_0^\infty \hat{G}_n \, d\hat{F}_n.$$

It is easily seen that $\hat{\alpha}_n > 0$ if $nC_n[x_{(i)}] > 1$ for all $i \le n - 1$; otherwise, $\hat{F}_n$ and $\hat{G}_n$ may be replaced by $F_n^\#$ and $G_n^\#$. Having estimated α, one may then estimate the population size by

$$\hat{N}_n = n/\hat{\alpha}_n.$$

EXAMPLE 1. When F and G are both the uniform distribution on the unit interval, $F_*(x) = x^2$ for $0 < x < 1$ and the conditional distribution of y_1 given x_1 is uniform on the interval $(0, x_1]$. To illustrate the properties of the estimators $\hat{F}_n$ and $\hat{G}_n$, $n = 10$ pairs of (x, y) values were simulated from the latter joint distribution. The results are listed in Table 1, along with the value of C_n and $\hat{F}_n$. Observe that there is only one data point in the interval $(0, 1/3]$ and four in the interval $(2/3, 1]$—reflecting the selection bias. The estimator $\hat{F}_n$ attempts to correct for this bias by assigning higher weight to the smaller values of $x_1, \cdots, x_n$. One may see the extent of this correction by comparing the observed average $\bar{x} = .612$ with the MLE of the mean of F, $\hat{\mu} = \int_0^1 x \, d\hat{F}_n = .519$. Of course, the means of F_* and F are $2/3$ and $1/2$. While assigning larger weights to smaller values may correct for some bias, it also increases variability. This is illustrated by the erratic behavior of $\hat{F}_n(x)$ for $x \le 1/2$.

4. Consistency. In this section, F and G denote continuous distribution functions for which $(F, G) \in \mathscr{H}$; and $(X_1, Y_1), (X_2, Y_2), \cdots$ denote i.i.d. random vectors for which $X_1 \sim F$ and $Y_1 \sim G$ are independent. We imagine the estimators

$\hat{F}_n$ and $\hat{G}_n$ computed from the populations $\mathscr{P} = \{1, 2, \cdots, N\}$ for $N = 1, 2, \cdots$ and investigate the limiting behavior of $\hat{F}_n$ and $\hat{G}_n$ as $N \to \infty$. Let (x_1, y_1), $(x_2, y_2), \cdots$ denote the successive values of (X_i, Y_i) for which $Y_i \le X_i$. Then $(x_1, y_1), (x_2, y_2), \cdots$ are i.i.d. with the common joint distribution function H_* of (2). As in Section 3, let $n = n_N = \#\{i \le N: Y_i \le X_i\}$ for $N \ge 1$. Then $n \sim$ Binomial(N, α) for all $N \ge 1$; and the conditional distribution of $(x_1, y_1), \cdots, (x_k, y_k)$ given $n = k$ is the same as their unconditional distribution for $1 \le k \le N$. Let P_n denote conditional probability given n. Below, the P_n-probability limits of $\hat{F}_n$ and $\hat{G}_n$ are determined as $n \to \infty$. It then follows that these are also the limits in unconditional probability as $N \to \infty$.

The following lemma may be of independent interest, since it computes the bias of the estimator $\hat{\Lambda}_n$.

LEMMA 2. *Suppose that F and G are continuous and that $(F, G) \in \mathscr{K}_0$. If h is a measurable function for which $\int_0^\infty |h| \, d\Lambda < \infty$, then*

$$E_n\left\{\int_0^\infty h \, d\hat{\Lambda}_n\right\} = \int_0^\infty h \, d\Lambda - \int_0^\infty h(1 - C)^n \, d\Lambda$$

for all $n \ge 1$, where $C(z) = \alpha^{-1}G(z)[1 - F(z)]$, $z \ge 0$. In particular,

$$E_n\{\hat{\Lambda}_n(x)\} = \Lambda(x) - \int_0^\infty (1 - C)^n \, d\Lambda, \quad 0 \le x < b_F, \ n \ge 1.$$

PROOF. If h is integrable with respect to Λ and $n \ge 1$, then

$$\int_0^\infty h \, d\Lambda_n = \textstyle\sum_{i=1}^n h(x_i)/nC_n(x_i).$$

Now, the conditional distribution of $nC_n(x_i) - 1 = \#\{j \le n: j \ne i, y_j \le x_i \le x_j\}$ given n and x_i is binomial with parameters $n - 1$ and $C(x_i)$ for each $i = 1, \cdots, n$. So,

$$E_n\{1/nC_n(x_i) \,|\, x_i\} = (1/nC(x_i))[1 - (1 - C(x_i))^n] \tag{11}$$

for all $i = 1, \cdots, n$, by an elementary calculation. Since $d\Lambda = dF_*/C$, the first assertion of the lemma now follows from multiplying (11) by $h(x_i)$, integrating over x_i, and summing over $i = 1, \cdots, n$. The second assertion then follows by letting h be the indicator of $[0, x]$ for fixed x, $0 < x \le b_F$.

Observe that the conditional bias of $\hat{\Lambda}_n(x)$ approaches zero as $n \to \infty$ for all $x < b_F$, but may do so arbitrarily slowly.

THEOREM 2. *Let F and G be continuous distribution functions for which $(F, G) \in \mathscr{K}$; and let F_0 and G_0 denote the conditional distributions of X_1 and Y_1 given $X_1 \ge a_G$ and $Y_1 \le b_F$, respectively. Then*

$$\sup_{x>0} |\hat{F}_n(x) - F_0(x)| \to 0 \leftarrow \sup_{y>0} |\hat{G}_n(y) - G_0(y)|$$

in P_n-probability, as $n \to \infty$.

PROOF. Since the distribution function H_* remains unchanged when F and

G are replaced by F_0 and G_0 by Lemma 1, it suffices to prove the theorem in the special case that $(F, G) \in \mathscr{H}_0$. Moreover, it suffices to prove the convergence of $\hat{F}_n$.

Given ε, $0 < \varepsilon < 1$, let $a > a_F$ be such that $\Lambda(a) < \varepsilon^2/4$ and let $B = B_{n,a}$ be the event $B = \{\hat{\Lambda}_n(a) \le \varepsilon/2\}$. Then

$$P_n(B') = P_n\{\hat{\Lambda}_n(a) > \varepsilon/2\} \le 2\varepsilon^{-1}E_n\{\hat{\Lambda}_n(a)\} \le \varepsilon/2$$

for all $n \ge 1$ by Lemma 2. So, since $\hat{F}_n(z) \le \hat{\Lambda}_n(a)$ and $F(z) \le \Lambda(a)$ for $z \le a$, it suffices to show that $P_n\{B, \sup_{x>a} | \hat{F}_n(x) - F(x) | \ge \varepsilon\} \to 0$ as $n \to \infty$.

Let $\lambda_{ni} = 1/nC_n(x_i)$ for $1 \le i \le n$; and define K_n and K by

$$K_n(x) = \prod_{i:a<x_i\le x} [1 - \lambda_{ni}]$$

and

$$K(x) = \exp\left\{-\int_a^x d\Lambda(z)\right\}$$

for $x \ge a$ and $n \ge 1$. Then $1 - \hat{F}_n(x) = [1 - \hat{F}_n(a)]K_n(x)$ and $1 - F(x) = [1 - F(a)]K(x)$ for all $x \ge a$ and $n \ge 1$. If B occurs, then

$$| \hat{F}_n(x) - F(x) | \le | K_n(x) - K(x) | + 3\varepsilon/4$$

for all $x \ge a$ and $n \ge 1$ by simple algebra. So, it suffices to show that $\sup_{x\ge a} | K_n(x) - K(x) | \to 0$ w.p.1 as $n \to \infty$ (on the space of (x_i, y_i), $i \ge 1$). In fact, since K is continuous and each K_n is monotone, it suffices to show that $K_n(x) \to K(x)$ w.p.1 for each fixed x, $a \le x < b_F$ (cf. Breiman, 1968, page 160).

Since $\sup_{x>0} | F_n^* - F_* | \to 0 \leftarrow \sup_{y>0} | G_n^*(y) - G_*(y) |$ w.p.1 as $n \to \infty$ and since C is positive and continuous on the interval (a_G, b_F), one finds that $\sup_{a\le z\le x} | 1/C_n(z) - 1/C(z) | \to 0$ w.p.1 as $n \to \infty$ for all x, $a < x < b_F$. So,

$$\hat{\Lambda}_n(x) - \hat{\Lambda}_n(a) = \int_a^x dF_n^*(z)/C_n(z) \to \int_a^x dF_*(z)/C(z) = \Lambda(x) - \Lambda(a)$$

w.p.1 as $n \to \infty$ for $a < x < b_F$. See Billingsley (1968, page 34). Since Λ is continuous and $\hat{\Lambda}_n$, $n \ge 1$, are monotone, the convergence must be uniform on $a \le x \le b$ for any $b < b_F$; and it follows that the maximum of λ_{ni} over any such interval $[a, b]$ approaches zero w.p.1 as $n \to \infty$. To complete the proof, let

$$R_n(a, x) = \sum_{i:a<x_i\le x} \log[1 - \lambda_{ni}] + [\hat{\Lambda}_n(x) - \hat{\Lambda}_n(a)] \tag{12}$$

for $a < x < b_F$ and $n \ge 1$. Then, by expanding $\log(1 - \lambda)$ in a Taylor series about $\lambda = 0$, one finds that there are intermediate points ξ_{ni} for which $| 1 - \xi_{ni} | \le \lambda_{ni}$ for $1 \le i \le n$,

$$| R_n(a, x) | = \tfrac{1}{2} \sum_{i:a<x_i\le x} \xi_{ni}^{-2}\lambda_{ni}^2 \to 0$$

and

$$K_n(x) = \exp\{-[\hat{\Lambda}_n(x) - \hat{\Lambda}_n(a)] + R_n(a, x)\}$$

$$\to \exp\{-[\Lambda(x) - \Lambda(a)]\} = K(x)$$

w.p.1 as $n \to \infty$ for $a < x < b_F$. This completes the proof.

COROLLARY 3. *If F and G are continuous and $(F, G) \in \mathscr{H}_0$, then* $\sup|\hat{F}_n - F| \rightarrow 0 \leftarrow \sup|\hat{G}_n - G|$ *in P_n-probability as $n \rightarrow \infty$.*

COROLLARY 4. *If F and G are continuous and $(F, G) \in \mathscr{H}_0$, then $\hat{\alpha}_n \rightarrow \alpha$ in P_n-probability as $n \rightarrow \infty$ and $\hat{N}_n/N \rightarrow 1$ in probability as $N \rightarrow \infty$.*

COROLLARY 5. *If F and G are continuous and $(F, G) \in \mathscr{H}_0$, then*

$$P_n\{nC_n[x_{(i)}] = 1, \text{ for some } i \leq n - 1\} \rightarrow 0$$

and

$$\min\{nC_n[x_{(i)}]: 1 \leq i \leq (1 - \varepsilon)n\} \rightarrow \infty$$

in P_n-probability as $n \rightarrow \infty$ for all ε, $0 < \varepsilon < 1$.

PROOFS. Corollary 3 is clear, and the convergence of $\hat{\alpha}_n$ to α in Corollary 4 follows. That $\hat{N}_n/N \rightarrow 1$ then follows, since $n/N \rightarrow \alpha$ w.p.1 as $N \rightarrow \infty$.

The second assertion in Corollary 5 follows from the relation

$$\hat{F}_n[x_{(i)}] - \hat{F}_n[x_{(i)} -] = \{1 - \hat{F}_n[x_{(i)} -]\}/nC_n[x_{(i)}]$$

for all $i \leq n$ and $n \geq 1$. Let $0 < \varepsilon < 1$ and $k = k(n, \varepsilon) = [(1 - \varepsilon)n] + 1$, where $[\cdot]$ denotes the greatest integer function. Then $1/\{1 - \hat{F}_n[x_{(k)}]\}$ is stochastically bounded and $\max_{i \leq n} \hat{F}_n(x_i) - \hat{F}_n(x_i -) \rightarrow 0$ in P_n-probability as $n \rightarrow \infty$, both by Theorem 2. This proves the second assertion in Corollary 4. The first assertion then follows from the second and its dual, obtained by reversing the roles of (X, Y) and $(1/Y, 1/X)$, by observing that $nC_n[x_{(i)}] = 1$ implies that $nC_n[y_{(i+1)} -] = 1$ for $1 \leq i \leq n - 1$.

REMARK 3. In the astronomy example, improved instrumentation might change m^*. In turn, this could change the definitions of Y, a_G, and F_0, the asymptotic value of $\hat{F}_n$.

REMARK 4. Since the joint distribution H_* depends on F and G only through F_0 and G_0, it is not possible to test the hypotheses $a_G \leq a_F$ and $b_G \leq b_F$ using $(x_1, y_1), \cdots, (x_n, y_n)$.

5. Convergence on compact intervals. For $0 \leq a < b \leq \infty$, let $\mathscr{D}[a, b]$ be the space of all functions f from $[a, b]$ into $R = (-\infty, \infty)$ which are right continuous on $[a, b)$, have left-hand limits on $(a, b]$, and are continuous at b. Endow $\mathscr{D}[a, b]$ with the Skorohod topology, as described by Billingsley (1968, Section 14). For each $n \geq 1$, define the stochastic processes X_n and Y_n by

$$X_n(t) = \sqrt{n}[F_n^*(t) - F_*(t)]$$

and

$$Y_n(t) = \sqrt{n}[G_n^*(t) - G_*(t)], \quad 0 \leq t < \infty.$$

where F_n^* and G_n^* are as in (5); and note the change in the use of the symbols

"X" and "Y." Then (X_n, Y_n) is a random element with values in $\mathscr{D}^2[0, \infty] = \mathscr{D}[0, \infty] \times \mathscr{D}[0, \infty]$ for each $n \ge 1$. If F and G are continuous, then the conditional distributions of (X_n, Y_n) given n converge

$$(X_n, Y_n) \Rightarrow (X, Y), \quad \text{as} \quad n \to \infty,$$

where X and Y are jointly Gaussian processes on $[0, \infty)$ with continuous sample paths and covariance structure

$$\begin{aligned} &\rho_{xx}(s, t) = F_*(s) - F_*(s)F_*(t), \qquad 0 \le s \le t < \infty, \\ (13) \qquad &\rho_{yy}(s, t) = G_*(s) - G_*(s)G_*(t), \qquad 0 \le s \le t < \infty, \\ \text{and} \qquad &\rho_{xy}(s, t) = H_*(s, t) - F_*(s)G_*(t), \quad 0 \le s, t \le \infty. \end{aligned}$$

Indeed, the convergence of the finite dimensional distributions of (X_n, Y_n) follows directly from the univariate central limit theorem and the Cramer-Wold device; and the tightness of the distributions of the pairs (X_n, Y_n), $n \ge 1$, follows from that of the components.

Observe that the covariance functions ρ_{xx}, ρ_{yy}, and ρ_{xy} may be consistently estimated.

Now suppose that F and G are continuous and that $(F, G) \in \mathscr{H}_0$. Fix values of a and b for which $a_G < a < b < b_F$ and let

$$\begin{aligned} W_{a,n}(t) &= \sqrt{n}\{[\hat{\Lambda}_n(t) - \Lambda(t)] - [\hat{\Lambda}_n(a) - \Lambda(a)]\} \\ (14) \qquad &= \int_a^t \frac{1}{CC_n}(X_n - Y_n)\, dF_n^* + \int_a^t \frac{1}{C^2} X_n\, dC \\ &\quad + X_n(t)/C(t) - X_n(a)/C(a) \end{aligned}$$

w.p.1 for $a \le t \le b$ and $n \ge 1$. The processes $W_{a,n}$, $n \ge 1$, are random elements with values in $\mathscr{D}[a, b]$.

THEOREM 3. *Suppose that F and G are continuous and that $(F, G) \in \mathscr{H}_0$. If $a_G < a < b < b_F$, then*

$$W_{a,n} \Rightarrow W^a = W_1^a + W_2^a, \quad \textit{as} \quad n \to \infty,$$

where

$$W_1^a(t) = \int_a^t C(s)^{-2}[X(s)\, dG_*(s) - Y(s)\, dF_*(s)]$$

and

$$W_2^a(t) = \frac{X(t)}{C(t)} - \frac{X(a)}{C(a)}, \quad a \le t \le b.$$

PROOF. First observe that $C = G_* - F_*$ is positive and continuous on $[a, b]$, since $a_G < a < b < b_F$. So, expressions like X/C and $\int_a^t C^{-2}X\, dG_*$ define continuous transformations from $\mathscr{D}[a, b]$ back into $\mathscr{D}[a, b]$. Since weak convergence is preserved by such continuous transformations, it suffices to show that

$$\Delta_n = \sup_{a \le t \le b} \left| \int_a^t \frac{1}{CC_n} Z_n \, dF_n^* - \int_a^t C^{-2} Z_n \, dF_* \right| \to 0$$

in P_n-probability as $n \to \infty$, with $Z_n = X_n - Y_n$, $n \ge 1$. To see this, one may replace C_n, F_n^*, and Z_n by other random elements, also denoted by C_n, F_n^*, and Z_n, which have the same joint distribution and converge to C, F_*, and $Z = X - Y$ w.p.1. as $n \to \infty$. See Skorohod (1956). That $\Delta_n \to 0$ w.p.1 then follows from Theorem 5.5 of Billingsley (1968) by considering a sequence t_n, $n \ge 1$, of random variables for which the supremum is nearly attained. The details are omitted. For a closely related argument, see Breslow and Crowley (1974, pages 447–448).

Of course, one would like to set $a = a_F$ in Theorem 3. If $a_G < a_F$, then this is possible. If $a_G = a_F$, then the limiting process may not be defined.

THEOREM 4. *Suppose that F and G are continuous, that $(F, G) \in \mathscr{H}_0$, and that $a_G = a_G$. If*

$$\int_{a_F}^{\infty} \frac{1}{G} \, dF < \infty \tag{15}$$

then $X(a)/C(a) \to 0$ and $W_1^a(t) \to \int_{a_F}^t C^{-2}[X \, dG_* - Y \, dF_*]$ in probability as $a \downarrow a_F$, for $a_F < t < b_F$. Conversely, if (15) fails, then the variance of $W_1^a(t)$ diverges to ∞ as $a \downarrow a_F$ for any $t \in (a_F, b_F)$.

PROOF. Recall that $C = \alpha^{-1}G(1 - F)$, so that $C(z) \sim \alpha^{-1}G(z)$ as $a \downarrow a_F$.

Suppose first that (15) holds. Then the variance of $X(a)/C(a)$ is at most $C(a)^{-2}F_*(a) \le [1 - F(a)]^{-2} \int_{a_F}^t (1/G) \, dF$, which tends to zero as $a \downarrow a_F$. Next, write $W_1^a = W_{11}^a - W_{12}^a$, where $W_{11}^a(t) = \int_a^t C^{-2} X \, dG_*$ and $W_{12}^a = \int_a^t C^{-2} Y \, dF_*$ for $a_F < a < t < b_F$. Thus, to show that $\lim W_1^a(t)$ exists in probability for all $t > a_F$, it suffices to show that the variances of $W_{11}^a(t)$ and $W_{12}^a(t)$ remain bounded as $a \downarrow a_F$ for some $t < a_F$. If $a_F < a < z < b_F$, then the variance of $W_{11}^a(z)$ is

$$\sigma_1^2(z) = 2 \int_a^z \int_a^t C(t)^{-2} C(s)^{-2} \rho_{xx}(s, t) \, dG_*(s) \, dG_*(t)$$

$$\le 2B \int_{a_F}^z \left[\int_s^z G(t)^{-2} \, dG(t) \right] G(s)^{-2} F_*(s) \, dG(s) \tag{16}$$

$$\le 2B \int_{a_F}^z G(s)^{-3} F_*(s) \, dG(s) \le 4\alpha^{-1} B \int_{a_F}^z (1/G) \, dF$$

for some constant B; and the last line is finite, by assumption. A similar argument shows that the variance $\sigma_2^2(z)$ of $W_{12}^a(z)$ remains bounded as $a \downarrow a_F$, if (15) holds.

If (15) fails, then a careful examination of (16) shows that $\sigma_1^2(z) \to \infty$

as $a \downarrow a_F$. $\sigma_2^2(z)$ may either diverge or remain bounded, depending on whether $\int_0^\infty (F/G)\,dF = \infty$ or $< \infty$, but one may show that $\sigma_2^2(z)/\sigma_1^2(z) \to 0$ in either case. That the variance of $W_1^a(z)$ diverges is an easy consequence. The details are omitted.

6. Convergence at an endpoint. In this section, we suppose that F and G are continuous, that $(F, G) \in \mathscr{H}_0$, and that (15) holds. In this case, the limiting distributions developed in the last section are valid when $a = a_G$. To avoid trivialities and simplify the notation, we suppose that $a_G = a_F = 0$ throughout. Fix a value of b for which $0 < b < b_F$ and define processes

$$W_n(t) = \sqrt{n}[\hat{\Lambda}_n(t) - \Lambda(t)]$$

and

$$Z_n(t) = \sqrt{n}[\hat{F}_n(t) - F(t)], \quad 0 \le t \le b,\ n \ge 1.$$

Then W_n and Z_n take values in $\mathscr{D}[0, b]$ w.p.1 for all $n \ge 1$.

THEOREM 5. *Suppose that F and G are continuous, that $(F, G) \in \mathscr{H}_0$, that* (15) *holds, and that $a_G = a_F = 0$. Then $W_n \Rightarrow W$ and $Z_n \Rightarrow Z$, as $n \to \infty$, where*

$$W(t) = \int_0^t C^{-2}[X\,dG_* - Y\,dF_*] + X(t)/C(t)$$

and

$$Z(t) = [1 - F(t)]W(t), \quad 0 \le t \le b,$$

with the convention $0/0 = 0$ when $t = 0$.

PROOF. That W is well defined follows from Theorem 4. To show that $W_n \Rightarrow W$ as $n \to \infty$, it suffices to show that $W_n(a) \to 0$ in P_n-probability as first $n \to \infty$ and then $a \to 0$. See Theorem 3. Now, as in (14),

$$W_n(a) = \int_0^a (1/CC_n)(X_n - Y_n)\,dF_n^* + \int_0^a C^{-1}\,dX_n$$

$$= I_n(a) + II_n(a), \quad \text{say},$$

for $a > 0$ and $n \ge 1$. Given $n \ge 1$, $II_n(a)$ is a normalized sum of i.i.d. random variables, and $E_n\{II_n(a)^2\} \le \int_0^a C^{-2}\,dF_*$ which is independent of n and tends to zero as $a \downarrow 0$, since $\int_0^b C^{-2}\,dF_*$ is finite. Thus, $II_n(a)$ converges to zero in P_n-probability as $n \to \infty$ and then $a \downarrow 0$. Next, recall that $d\hat{\Lambda}_n = dF_n^*/C_n$ and write

$$|I_n(a)| \le \int_0^a C^{-1}|X_n - Y_n|\,d\hat{\Lambda}_n \le B_{n,a}\int_0^a C^{-1}\,d\hat{\Lambda}_n$$

for $a > 0$ and $n \ge 1$, where $B_{n,a} = \sup_{t \le a}|X_n(t) - Y_n(t)|$. Now $B_{n,a} \to 0$ in

P_n-probability as $n \to \infty$ and then $a \downarrow 0$; and, by Lemma 2, $E_n\{\int_0 C^{-1}\, d\hat{\Lambda}_n\} \le \int_0^a C^{-1}\, d\Lambda$, which is independent of n and tends to zero as $a \downarrow 0$. This completes the proof that $W_n \Rightarrow W$ and $n \to \infty$.

Now consider Z_n. With R_n as in (12),

$$Z_n(t) = [1 - \hat{F}_n(t)]\,\sqrt{n}\left\{\exp\left[\frac{1}{\sqrt{n}}\, W_n(t) - R_n(0,\, t)\right] - 1\right\} \tag{17}$$

for $0 \le t \le b$ and $n \ge 1$. So, it suffices to show that $\max_{t \le b} \sqrt{n}\,|\,R_n(0,\, t)\,| \to 0$ in P_n-probability as $n \to \infty$. Now, $\max_{t\le b}\,|\,R_n(0,\, t)\,| = |\,R_n(0,\, b)\,|$; and

$$|\,R_n(0,\, b)\,| \le B_n \textstyle\sum_{i:x_i \le b} 1/\{nC_n(x_i)[nC_n(x_i) + 1]\}, \tag{18}$$

where $B_n = \max\{\xi_{ni}^{-2}\colon x_i \le b\}$ and ξ_{ni}, $1 \le i \le n$, are intermediate points as in (12). Now, B_n is bounded in P_n-probability, by Corollary 3; and the expectation of the sum in (17) is at most $(1/n)\int_0^b C^{-2}\, dF_*$, as in the proof of Lemma 2. (The conditional distribution of $nC_n(x_i) - 1$ given n and x_i is binomial $[n - 1,\, C(x_i)]$ for $1 \le i \le n$.) Thus, $R_n(0,\, b) = O_p(1/n) = o_p(1/\sqrt{n})$ in P_n-probability to complete the proof.

REMARKS 5. By Corollary 5, Theorems 2, 3, and 5 are valid if $\hat{F}_n$ is replaced by the modification $F_n^{\#}$ of (9), provided the constants $k_{n1}, \ldots, k_{nn}$ are bounded. Indeed, Corollary 5 asserts that $P_n\{F_n^{\#}(z) = F_n(z) \text{ for all } z \le b\} \to 1$ as $n \to \infty$ for any $b < b_F$, in this case.

6. There is a dual to Theorem 5. Suppose that F and G are continuous, that $(F,\, G) \in \mathscr{K}_0$, that $b_G = b_F = \infty$, and that $\int_0^\infty 1/(1 - F)\, dG < \infty$. Let $U_n(t) = \sqrt{n}[\hat{G}_n(t) - G(t)]$, $t \ge 0$, $n \ge 1$, and regard U_n as random elements with values in $\mathscr{D}[a,\, \infty]$, where $a > a_F \ge a_G$. Then $U_n \Rightarrow U$, where

$$U(t) = -G(t)\left\{\int_t^\infty C^{-2}(X\, dG_* - Y\, dF_*) - Y(t)/C(t)\right\}$$

for $a \le t < \infty$.

7. The condition (15) is not surprising, since it is necessary for the convergence in distribution of $\sqrt{n} \times$ estimation error even in the case when G is known. In this case, the nonparametric maximum likelihood estimator of F is $F_n(t) = [\sum_{i:x_i\le t} 1/G(x_i)]/[\sum_{i=1}^n 1/G(x_i)]$ for $t \ge 0$ and $n \ge 1$; and it is easily seen that $\sqrt{n}[F_n(t) - F(t)]$ converges in distribution for all $t > a_F \ge a_G$ iff (15) holds.

8. If (15) fails, then other limiting distributions may obtain. Suppose, for example, that F is continuous, that $a_F = 0$, and that $G = F^c$, where $1 < c < \infty$. Let $\delta = 1/(1 + c)$. Then $n^\delta[F_n(t) - F(t)]$ has a limiting stable distribution for all $t > 0$ for which $F(t) < 1$. To see this, fix t and write

$$\hat{\Lambda}_n(t) = \textstyle\sum_{i:x_i\le t} \dfrac{1}{nC(x_i)} + \displaystyle\int_0^t \left[1 - \frac{C_n}{C}\right] d\hat{\Lambda}_n = I_n + II_n.$$

Then $|\,II_n\,| \le \max_{s\le t}\,|\,C_n(s) - C(s)\,|\,\int_0^t (1/C)\,d\hat{\Lambda}_n = O_p(1/\sqrt{n})$ by Lemma 1 and properties of empirical processes. Let $z_i = (1/C(x_i))I_{(0,t)}(x_i)$ for $i = 1, 2, \cdots$. Then $z_i, z_2, \cdots$ are i.i.d. with common mean $\Lambda(t)$. Now, it is easily seen that z_i is in the domain of attraction of a stable distribution with characteristic exponent $\gamma = (1 + c)/c$ and skewness parameter 1, in Feller's (1966, pages 540–543) terminology. So, $n^\delta[\hat{\Lambda}_n(t) - \Lambda(t)]$ has a limiting stable distribution as $n \to \infty$. (In fact, the same stable distribution is obtained for all t.) That $n^\delta[\hat{F}_n(t) - F(t)]$ has a limiting distribution, now follows from (17) by using a stable distribution to bound $z_1^2 + \cdots + z_n^2$ in (18).

Acknowledgements. Irving Segal introduced me to this problem through the application to astronomy. Gary Lorden brought the papers by Lynden-Bell (1971) and Jackson (1974) to my attention. The referees and associate editor contributed useful comments and criticisms.

REFERENCES

BAROUCH, E. and KAUFMAN, G. M. (1975). Probabilistic modelling of oil and gas discovery. *Energy* 133–152.

BHATTACHARYA, P. K. (1983). Justification for a K-S type test for the slope of a truncated regression. *Ann. Statist.* **11** 697–701.

BHATTACHARYA, P. K., CHERNOFF, H. and YANG, S. S. (1983). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.* **11** 505–514.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading, Massachusetts.

BRESLOW, N. and CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2** 437–453.

FELLER, W. (1966). *An Introduction to Probability Theory and its Applications, Vol. 2.* Wiley, New York.

JACKSON, J. C. (1974). The analysis of quasar samples. *Mon. Not. R. Astr. Soc.* **166** 281–295.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Soc.* **53** 457–481.

KRAMER, M. (1983). Stopping a size dependent exploration process. Ph.D. Thesis, The University of Michigan.

LYNDEN-BELL, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Mon. Not. R. Astr. Soc.* **155** 95–118.

NICOLL, J. F. and SEGAL, I. E. (1980). Nonparametric elimination of the observational cutoff bias. *Astron. Astrophys.* **82** L3–L6.

SEGAL, I. E. (1975). Observational validation of the chronometric cosmology: I. Preliminaries and the red shift-magnitude relation. *Proc. Nat. Acad. Sci.* **72** 2437–2477.

SKOROHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probab. Appl.* **1** 261–290.

STARR, N. (1974). Optimal and adaptive stopping based on capture times. *J. Appl. Probab.* **11** 294–301.

STARR, N., WARDROP, R. and WOODROOFE, M. (1976). Estimating the mean from delayed observations. *Z. Wahrsch. verw Gebiete* **35** 103–113.

VARDI, Y. (1982a). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10** 616–620.

VARDI, Y. (1982b). Nonparametric estimation in renewal processes. *Ann. Statist.* **10** 772–785.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109